

# Exhibit 7

**THE UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF COLUMBIA**

STATE OF TEXAS,

*Plaintiff*

vs.

ERIC H. HOLDER, JR.,  
in His Official Capacity as Attorney  
General of the United States,

*Defendant.*

Case No. 1:12-CV-00128

(RMC, DST, RLW)  
Three-Judge Court

**SUPPLEMENTAL EXPERT DECLARATION OF THOMAS SAGER**

1. I have reviewed the report of the Department of Justice's expert, Dr. Stephen Ansolabehere ("SA").

2. The State has requested that I provide additional analysis of the work of SA. In general my critique of SA's work is limited to his "Protocol for Matching Databases" that appears at ¶¶19-29 of his report. Several of SA's opinions involve qualitative assessments about who is or is not more likely to be affected by SB 14. I note those qualitative assessments, but it is not within the scope of my first declaration nor this supplemental declaration to opine as to their validity.

3. Specifically, I was asked to undertake two additional tasks:  
(1) I was asked to match SA's VRNID dataset to the State Driver License ("DL") and License to Carry ("LTC") datasets using alternative matching

criteria, as well as to screen for “age over 65” and “voter registration suspense” statuses. Both the State and SA derived lists of registered voters that lack apparent matches to the DL and LTC datasets after application of their matching criteria. In my earlier declaration, I discussed the results of my attempts to find additional matches for the State’s list (“May No Match”). I am now asked to attempt to find additional matches for SA’s list (“VRNID”). (2) I was asked to consider if there might be statistical and/or data processing reasons that might account for the much larger size of SA’s no-match list (VRNID, with 1,501,977 ultimate entries) compared with the State’s May No Match list (with 588,095 entries).

4. I understand that SA cleaned both the State voter registration (“VR”) and DL/LTC databases before attempting to match them. SA reports that he found 13,072,454 records in the original VR database and that he removed 273 of these records because of duplicate voter ID numbers, 25 because of duplicate SSNs and birth dates, and 6,652 because of duplicate first name, last name, date of birth and street address (SA ¶ 19). SA also reports that he found 125,015 VR records lacking SSNs and having very common names and that he removed these records for most of his analysis (SA ¶ 22).<sup>1</sup>

5. SA reports that he found 25,985,555 records in the original DL database and that he removed numerous records for the reasons listed in the

---

<sup>1</sup> In SA ¶ 26, SA refers to these 125,015 records as being in the DL database. I assume he means they are in the VR database.

following table (SA ¶ 20). This left 19,951,173 records in the DL database.<sup>2</sup>

He also reports that there are 266,151 DL records with “ambiguous status” that he both included and excluded in two separate runs of his analysis (SA ¶ 22).

#### SA Deletions from State DL Database Prior to his Matching Sweeps

A	287,236	no drivers license or State ID
B	3,144,900	drivers license expired for more than 2 years
C	1,535,504	drivers license expired between 60 days and 2 years
D	779,918	deceased driver
E	724,974	duplicate SSN
	<b>6,472,532</b>	<b>TOTAL</b>

6. SA also reports that he found 840,664 records in the original LTC database and that he removed numerous records for the reasons listed in the following table (SA ¶23). This left 592,270 records in the LTC database.<sup>3</sup>

#### SA Deletions from State LTC Database Prior to his Matching Sweeps

2,338	Deceased
1	Unreadable
38,919	Non-U.S. citizen
12,437	failed application
194,669	nonrenewable expired license
<b>248,364</b>	<b>TOTAL</b>

---

<sup>2</sup> Apparently, the deletions described by SA in SA ¶ 20 are not mutually exclusive, for the total deletions that he enumerates would leave  $25,985,555 - 6,472,532 = 19,513,023$  records instead of the 19,951,173 that he asserts.

<sup>3</sup> Subtraction yields  $840,664 - 248,364 = 592,300$  remaining records – a small discrepancy from the 592,270 reported by SA.

7. With the thus-cleaned databases, SA conducted three consecutive and apparently cumulative sweeps to find matches for VR records in the DL and LTC databases based on the following matching criteria, in the order given:

(Sweep #1) identical SSN

(Sweep #2) identical date-of-birth and identical first name and identical last name (all three criteria required for a match)

(Sweep #3) identical date-of-birth and identical first name and identical last name and identical middle name (all four criteria apparently required for a match)

8. As a side note, it is not clear whether SA's sweep #3 will pick up any additional records. Any records in VR that match DL/LTC on the basis of DOB+FN+LN+MN in sweep #3 will have already been matched on the basis of DOB+FN+LN in sweep #2.

9. As a result of application of his three sweeps, SA matched all but 1,893,143 VR records to DL/LTC (SA ¶ 26). This count includes 125,015 with "insufficient information to match" and 266,151 with "ambiguous DL status" (SA ¶ 26). These 1,893,143 "no match" VR cases constitute SA's VRNID list. It is this VRNID database (reduced by exclusion of the 125,015 and 266,151 aforementioned records) that the State has asked me to try to match to the DL/LTC databases.

10. SA has provided his VRNID database to the State (absent sensitive SSN data). The State was able to resupply the missing SSN information with high confidence by matching on the unique voter ID field. Therefore, the VRNID was available to me for various types of matching strategies. After matching VRNID to DL/LTC, I found that most of the records in VRNID either match DL/LTC or fall into special categories like “over age 65” that I understand are significant for the purpose of SB 14.<sup>4</sup>

11. Before I discuss my attempts to match SA’s VRNID to DL/LTC, I will discuss a number of obvious problems with the data cleaning decisions that SA made that have the effect of inflating the number of entries in his VRNID.

12. First, SA removed 779,918 deceased drivers from DL prior to his matching sweeps, but he did not remove them from VR (see category D in my ¶5). Therefore, any deceased drivers who are in the VR database will remain unmatched and will end up in VRNID. Because of this, there are 57,718 deceased persons in SA’s VRNID. Presumably, the dead do not vote and therefore should not be included on a list of those potentially affected by SB 14. SA could have avoided this problem by leaving deceased drivers in the

---

<sup>4</sup> Because I was out of the state during the period of this analysis and because of the sensitivity of social security information, the analysis was performed by a technician at the Office of the Attorney General Legal Technical Support division under my remote supervision and direction. I reviewed and quality checked all of the matches by receiving random samples and reviewing them for accuracy.

DL database. By that means, deceased (and formerly driving) voters could have been matched and removed and not end up in VRNID.

13. Second, SA removed 4,680,404 records from DL prior to his matching sweeps on account of expired drivers licenses (categories B and C in my ¶ 5). These removals are problematic for a number of reasons:

14. (i) Some expired licenses no doubt belong to voters who have moved out of State and therefore are no longer eligible to vote in Texas. Such (former) voters will end up in VRNID, although they – like deceased voters – presumably should not be counted among the unmatched.

15. (ii) I understand that registered Texas voters who are over age 65 are automatically entitled to vote by mail and disabled voters are exempt from photo ID requirements. I understand that these exemptions apply regardless of the expiration status of a voter's driver license. No doubt many voters holding expired licenses are over 65 or disabled. Deletion of expired drivers licenses from DL would be unexceptionable for the purpose of age-over-65 determination, provided an additional screen of the VR database for age were added.<sup>5</sup> Disabled voters with expired licenses would end up in VRNID without the capacity to detect their exemption from SB 14 on account of disability.

16. (iii) I understand that drivers holding licenses that are expired less than two years may conveniently renew online, and many in this group

---

<sup>5</sup> SA did not screen for either disability or age over 65.

may choose to do so. It is clear that the effect of the three factors listed above is to inflate the size of SA's VRNID.

17. Third, SA removed 724,974 records from DL prior to his matching sweeps on account of duplicate SSNs (category E in my ¶ 5). These removals also are problematic. They do not affect SA's sweep #1 – on SSN – but they do affect his sweep #2 and sweep #3. The reason is subtle but clear. The following is an illustrative hypothetical example. Suppose that I am registered to vote as Tom Sager with SSN=3333 and correct birth date.<sup>6</sup> Suppose further that I have two DL entries – one as Tom Sager with SSN=111-22-3333 and the other as Thomas Sager with SSN=111-22-3333, and both have my correct birth date. SA's procedure would delete one of my two DL entries before matching. Whichever one he deletes, he will not match me on his sweep #1 and will need to proceed to sweeps #2 and #3. If he deletes the Tom Sager DL entry, then he will fail to match me on sweeps #2 (DOB+FN+LN) and #3 (DOB+FN+LN+MN) and will put me into VRNID. Only if he deletes the Thomas Sager DL entry will he successfully match me and keep me out of VRNID. In other words, it is not necessary to delete duplicate SSNs from DL/LTC prior to matching sweeps. Retention of SSN duplicates increases the chance of matching voters to true variant names and spellings. Deletion of SSN duplicates tends to inflate the size of VRNID.

---

<sup>6</sup> 54.6% of the VR database lack full SSN.



18. There remains one SA deletion category (A) referenced in my ¶5 that I have not yet discussed. Category A numbers 287,236 DL entries that SA removed because no license or ID was shown as issued. These removals are unobjectionable on their own but result in no major difference. I inquired whether the State had deleted Category A from the DL database for the May No Match dataset of 588,095 voters. Since it had not, on this account the size of the May No Match database was slightly smaller than it should have been. However, the difference was small – only 8,228 records.

19. SA also made a deletion from the LTC database that will tend to inflate his results. SA deleted 38,919 records because they were recorded as “non-citizens” in the LTC database. Non-citizens are not eligible to vote, and so if these records matched to a VR entry that would either indicate: (a) an ineligible voter that should not be counted as a bona fide voter without ID or (b) a voter who had subsequently naturalized and had an LTC ID and therefore should not be counted as a voter without ID. Either way, SA should not have eliminated these entries. However, the LTC dataset produces so few additional matches that the elimination of this small number of records likely is not nearly as significant as the SA’s other deletions of deceased and expired records from the DL dataset.

20. I now discuss the protocols for and results of my attempts to match SA’s VRNID database to the DL/LTC databases. In spite of the problems with the construction of SA’s VRNID database that I discussed in

the preceding paragraphs, I took SA's VRNID database as he provided it for the purpose of my assignment to match it to the DL/LTC databases. Except that, for this purpose, the VRNID database to match consists of the 1,893,143 "no match" VR cases identified by SA, reduced by exclusion of the 125,015 and 266,151 VR records that SA identified as having "insufficient information" or "ambiguous DL status" (see my ¶ 9).

21. Working with this VRNID database of 1,501,977 voter records, I supervised a re-match to the full DL/LTC databases as they were prior to SA's cleaning.<sup>7</sup> Performing this re-match required first re-appending the SSN data that SA had redacted from the version of VRNID that he delivered to the State.

22. The following four match sweeps were run to match SA's VRNID to both the DL and LTC databases. In each sweep, identical matches on all criteria listed were required.

- a. First Name and Last Name and DOB
- b. SSN9
- c. First Name and Last Name and SSN4
- d. SSN4 and DOB

---

<sup>7</sup> As I indicated above, I remotely supervised OAGLTS personnel in the matching since I was out of the state.

23. In the same manner as described in my June 1, 2012 declaration, I also identified “suspense” entries and over-age-65<sup>8</sup> entries for SA’s VRNID database, as I had done for the State’s May No Match database.

24. Finally, as a check on SA’s cleaning of the DL database, I directed a separate match of VRNID to only those DL records that SA cleaned from the DL database on account of no ID or out-of-state residence, deceased drivers, or expired licenses. The purpose of this separate and more restricted sweep is to discover the matches that SA failed to find as a result of his exclusion of these categories of DL entries. For this separate sweep, all four of the matching criteria at my ¶ 22 were employed. I first discuss the results of this check on the effects of SA’s cleaning of the DL database.

25. First, only 746 of VRNID match a no-ID DL entry.

26. Second, SA’s removal of DL records for deceased drivers but not for deceased voters makes a more substantial difference. 57,718 of VRNID match DL entries for deceased drivers.

27. Third, 468,775 records in VRNID match a DL record with an expired license.

28. The following table shows a summary of the results of matching all of VRNID to the records that SA cleaned from DL by cleaning category matched.

---

<sup>8</sup> Over age 65 as of November 6, 2012.

	DPS_EXPIRED	DPS_OTHER_JURIS	DPS_DECEASED	DPS_Nold	Count
	1	1			102,951
	1		1		28,693
	1				337,131
		1			10,320
			1		29,025
				1	746
<b>Not in a "cleaning" category</b>					<b>993,111</b>
<b>Total</b>					<b>1,501,977</b>

29. The following table expands the preceding table by adding a breakdown by over-age-65 status and by "suspense" status. I understand that voters over the age of 65 are not required to present a photo id in order to vote under SB 14 because they can vote by mail. I further understand that a voter registration in "suspense" means that mail sent to the voter has been returned to sender and that the voter has not yet verified a new address, meaning they may have moved out of state or otherwise become ineligible. Although one can tally from this table 468,775 records in VRNID that match a DL record with an expired license, 110,073 of these are for voters over the age of 65, and 138,426 of these entries are in suspense status.

65	Suspense	DPS_EXPIRED	DPS_OTHER_JURIS	DPS_DECEASED	DPS_Nold	Count
1	1	1	1			6,318
1	1	1		1		5,999
1	1	1				17,320
1	1		1			196
1	1			1		5,095
1	1				1	4
1	1					25,792
1		1	1			5,019
1		1		1		12,733
1		1				62,684
1			1			758
1				1		14,048
1					1	36
1						174,415
	1	1	1			41,492
	1	1		1		3,064
	1	1				64,233
	1		1			2,430
	1			1		2,873
	1				1	156
	1					161,127
		1	1			50,122
		1		1		6,897
		1				192,894
			1			6,936
				1		7,009
					1	550
<b>Not in a "cleaning" category or suspense or over 65</b>						<b>631,777</b>
<b>Total</b>						<b>1,501,977</b>

30. I turn now to the results of matching VRNID to DL/LTC using the four matching sweeps outlined in my ¶ 22. Exhibit A to this declaration shows the results of this re-match analysis, with separate break-outs for each of the four sweep criteria applied, as well as the overlaps among the four

sweep criteria, and a further breakdown for over age 65 status and suspense status. A partial summary of these results is shown in the table below.

Criteria	Number of VRNID entries
Matched to DPS DL or LTC using any criteria from ¶ 22 (a) – (d)	814,903
Over 65	330,377
Suspense	335,939
Reported ID Number to SOS but did not match any of above	261,887

31. By summing all rows in Exhibit A, one can readily calculate that out of 1,501,977 entries in SA's no-match VRNID database, a total of 1,072,366 voter entries either match one or more of the four matching sweeps in my ¶ 22 (a) – (d), or are over 65 or are in suspense. Furthermore, by summing appropriate rows in Exhibit A, one can calculate that 814,903 of these VRNID entries qualify as matches by application of one or more of my four matching sweeps. This count includes 210,601 matches on full SSN9. By summing appropriate remaining rows, one computes that an additional 141,182 voter entries are over 65, and the balance of 117,094 voter entries are in suspense. That leaves a remainder of 429,611 ( $= 1,501,977 - 1,072,366$ ) as yet unaccounted for. This number includes 746 VRNID entries that match to the no-ID category in the DL database. An additional 261,887 are voter registrants who reported no official ID number to the Texas Secretary of State. This leaves only 167,724 of VRNID unaccounted for. Many of these entries (and the larger set of approximately 700,000 entries that do not formally "match" into the DPS database using the criteria of ¶ 22

(a)-(d)) likely do have undetected matches in the DPS database. Over 50% of the VR database lacks full SSN data, but the discussion below and in Exhibit B show that there is a very high rate of name and DOB mis-matches for SSN matching records through alternate names, marriage, or data entry errors.

32. As noted in the table above (my ¶ 30), 261,887 VR entries lack an official ID number. I understand that when a voter registers, the Texas Secretary of State (“TXSOS”) records the voter’s assertion that he/she has state identification. I understand that when in the creation of the “No Match” sets that I discussed in my first declaration, any voter registration entry that includes state ID as belonging to someone was removed from the No Match set. Thus, TXSOS generally took people at their words that they had driver licenses or state IDs when they registered. These individuals (those who did not otherwise match) are listed in the last row of my matching table (my ¶ 30), above, and represent a significant number.

33. As I also noted, 167,724 entries in SA’s VRNID database remain unaccounted for by the above procedures. Of these remaining 167,724 voter entries, 31.6% have Spanish surnames, per the appropriate VR data field. This rate is 9.3% higher than the overall Spanish Surname registration rate of 22.25% reported by SA (SA ¶ 33). The difference represents less than 16,000 ( $9.3\% \times 167,724$ ) of the unaccounted for voters in VRNID. 16,000 represents approximately 0.55% of the total Spanish surname population of registered voters. On general statistical and data processing grounds, there

are eminently reasonable bases for doubting the meaningfulness of such a relatively small difference. Although such a difference could be “statistically significant” by the rote application of a standard statistical test, neither the statistical grounds for use of such a test nor the justification for interpreting its meaning as implying lesser Hispanic access to State identification has been established to a reasonable certainty, in my opinion.

34. For example, if Hispanics have even a slightly higher rate than non-Hispanics for holding the appropriate federally issued identification, the ethnic difference could be mooted. Also, uncertainties about the quality of data (about which I have much more to say below) could turn 16,000 into a rounding error. If Hispanics have even a slightly higher probability than non-Hispanics of name variants or misspellings, SA’s match sweeps will tend to place relatively more of them into his unmatched VRNID database.

35. The operation of such a selection bias can be demonstrated for females in SA’s VRNID database. Females appear to be over-represented in VRNID. SA’s sweep matching criteria are less likely to match females than males. One reason is that many females change their last names when they marry and therefore are more likely to have different last names in VR than in DL/LTC. SA’s matching criteria that require identical first and last names will assign such females incorrectly to VRNID. The effect can be substantial. In an empirical analysis of VRNID that I discuss in greater detail in Exhibit C. I took a random sample of 1,000 entries from a subset of VRNID known to



have matches in DL/LTC with very high confidence. 650 of the 1,000 are female and only 350 are male. Such a disparity cannot be explained by chance. I counted approximately 207 females who appear to have changed their last names. Further analysis in Exhibit C confirms that in matching criteria that do not require first and last name matching, there is a selection bias for male matches.

36. As a statistician, I often perform statistical tests of the difference between two percentages. Those tests are designed to determine whether a difference is real and not attributable to chance; the tests are not designed to determine whether the difference is meaningful, nor to determine the cause of the difference. If one has enough data, nearly every difference will test as real. Meaningfulness must be judged by other criteria. Moreover, the validity of such tests depends upon satisfying the assumptions upon which they depend. Furthermore, if one is not careful, then the cause of a statistically significant difference may not be what one assumes. For example, a higher rate of Spanish than non-Spanish surnames among unmatched voters may not result from less access to state identification by those with Spanish surnames but may result from a higher rate of data errors in Spanish surnames. Above, I cite the over-representation of females in VRNID relative to males. Does this imply that Texas females have less access to State identification than males? No, it just means that there are more “data errors” for females in the voter and driver license databases on

account of many Texas females following the social custom of adopting their husbands' last names in marriage.

37. In the current case, both SA and I have reported most of our tallies with a precision that belies potentially large uncertainties about the quality of the data that we used and limitations in the available means for measurement. I think this point about data uncertainties is quite important. In fact, I have conducted an investigation into the quality of some of the data in VRNID. I will briefly mention five data problems here, but leave the detailed discussion to Exhibit B.

38. Name variants. Many names have variants and nicknames. "Thomas", "Tom", "Tommy" are all variants of the same name. The requirement of an identical match on name will fail to match if the VR database has a different variant from the DL/LTC database.

39. Name misspellings. A clerical error in reading or typing a name can result in a misspelling and consequent mismatch. The requirement of an identical match on name will fail to match me if I am in the VR as "Tomas" and in DL as "Thomas".

40. Married female name changes. If my wife had a driver's license before she married me and registered to vote after marriage and changed her last name after marriage, then she would be in DL as "Alexander" and in VR as "Sager". Consequently, the requirement of an identical match on name will fail to match her.

41. Date of Birth. Clerical errors in reading and/or data input may result in dates of birth that are slightly different between VR and DL/LTC. For example, a hand-written date of “4/19/46” might be misread as “4/14/46” if the “9” is misread as a “4”.

42. Differential name error rates by ethnicity. To a lay mind, it is possible to think of reasons why Spanish surnames may be more prone to data entry error than non-Spanish surnames. For example, a data entry clerk may not be familiar with Spanish surnames and perhaps be more likely to misspell or misinterpret them. For example, is a hand-written “DE LA CRUZ” one word, two words, or three words – or should it be “DE LA CRUS?”

43. Exhibit B discusses a quantitative analysis of data issues in SA’s no-match VRNID by using a random sample of 1000 entries taken from a subpopulation of VRNID entries that are known with high confidence to have been correctly matched to the same people in DL/LTC. Since this sample of 1000 people have been correctly matched, their first names, last names, and dates of birth should be the same in VR and in DL/LTC. Yet many are not. 340 fail to match on exact first name, 388 fail to match on exact last name, and 368 fail to match on exact date of birth. To be sure, such high error rates should not be extrapolated to the VR and DL/LTC databases generally, since the VRNID database was formed from VR entries whom SA had difficulty matching on these criteria. However, this sample of 1,000 known true matches provides a laboratory for testing the prevalence of

VRNID data issues without the otherwise legitimate concern that differences in names and DOBs might be true differences resulting from the mismatch of different people. These are the same people, yet their VR data is often not the same as their DL/LTC data. The effect of data errors in first name, last name, and/or date of birth is to reduce the chance of finding true matches and hence to inflate the size of VRNID.

44. Exhibit D discusses the expected matching error rates on the matching criteria other than SSN9.

I swear the foregoing is true and correct to the best of my knowledge.

Dated: June 11, 2012

Respectfully submitted.

Thomas W. Ager

Kihei, Hawaii